Genet Program Evolvable Mach (2008) 9:171–182
DOI 10.1007/s10710-008-9060-3

ORIGINAL PAPER

An eigen analysis of the GP community

W. B. Langdon · R. Poli · W. Banzhaf

Received: 24 October 2006 / Revised: 25 May 2007 Published online: 15 April 2008

Abstract The coauthorship and coeditorship relations as recorded in the genetic programming bibliography provide a quantitative view of the GP community. Eigen analysis is used to find the principle components of the community. It shows the major eigenvalues and eigenvectors are responsible for 70% of the connection graph. Top eigen authors are given.

Keywords Genetic programming · Bibliography · Social network · Eigen analysis

1 Introduction

In recent years, bibliometric analysis of scientific texts has become a topic of research interest in its own right. Although it was realized early on that scientific productivity could indeed be scrutinised by scientific methods [1, 2], it is only with the recent advent of online bibliographies and text repositories, with the enormous gain in computational power afforded by new generations of computing machines, and with the improved sophistication of algorithms, that the field has become accessible for in-depth analysis.

W. B. Langdon
Departments of Biological and Mathematical Sciences, University of Essex, Colchester, UK
URL: http://www.cs.essex.ac.uk/staff/W.Langdon/

R. Poli (✉)
Department of Computing and Electronic Systems, University of Essex, Colchester, UK
e-mail: rpoli@essex.ac.uk
URL: http://dces.essex.ac.uk/staff/rpoli/

W. Banzhaf
Department of Computer Science, Memorial University of Newfoundland, St.John's, Canada
URL: http://www.cs.mun.ca/~banzhaf/

Based on the idea that science is performed as a social activity, the notion of social networks [3] has played an ever increasing role in the analysis of developments in science [4, 5]. Co-authorship networks are particularly interesting, as they reveal patterns of activity that both relate to single individuals and to the development of an entire field.

There are various tools for analysis, among them statistical analysis, visualisation, eigen analysis, and temporal analysis of these networks. We shall concentrate on eigen analysis.

Expressed as matrices, coauthorship networks are amenable to quantitative algebraic analysis. Eigen analysis can be thought of as similar to the way that the mean, variance, skewness, kurtosis and higher order cumulants can be used to summarise a distribution. Given enough higher order cumulants the complete distribution can be recovered exactly but the lower order cumulants (mean, variance,...) are often used to summarise it. In the same way the complete set of eigenvectors and their eigenvalues fully describes a matrix but the principle modes of the matrix are summarised by the eigenvectors with the largest (absolute value) eigenvalues. Thus the first eigenface captures the similarity between people's faces, and higher order eigenfaces describe in increasing detail variation between faces [6, 7]. Although linear algebra is a classic technique, there have been interesting new uses for eigenvectors and their associated eigenvalues. Even the career choices of medical students [8] have been subjected to this kind of examination. In his wide raging analysis of co-citations of 50 leading medical informaticians [9], Andrews includes eigenvectors to capture the variability in his data.

In contrast to that study of co-citations among a small number of established experts we will use eigenvalue analysis of coauthorship on the whole of the genetic programming field as captured by the GP Bibliography http://www.cs.bham.ac.uk/~wbl/biblio/. We use the same version of the GP-bibliography as was recently analysed by Tomassini and colleagues [10].

2 Co-authors in the GP bibliography

As of April 2006, the bibliography contained information on 3 078 GP authors and 4,128 papers, books, proceedings, technical reports etc. (See [11] for a recent study of the GP literature.) We concentrate upon the relationships between authors held in the bibliography. In particular we look at the 11 005 links between authors due to two or more people collaborating on writing (or editing) an entry in the bibliography. These links form a co-authorship graph.

As with other fields, the GP field is not fully interconnected by joint publications. That is the whole co-authorship graph is not a "small world" network, although Tomassini et al. [10] show that in important respects the connected component behaves like a small world network. In restricted domains, like literature studies, it is common for the graph to fall into disconnected parts. This is because the social interactions we look at, joint publications, are sparsely distributed. That is no one has published papers with more than a tiny fraction of other people in their field. This is certainly true of genetic programming. However, the data is both available

and quantitative. In contrast, the criterion used the original small world study [12] was "to be on a first name basis" which is less tangible though there are many such links. Having many links virtually guarantees the graph is connected. A connected graph means the giant component encompasses everyone.

Although it would be possible, with care, to conduct an eigen analysis of the complete coauthorship graph, we shall concentrate on the largest connected component, since it is of the most general interest. This component contains 942 people who are together responsible for 2,144 entries, see Fig. 1. We used Graphviz' neato to layout the graph. Neato uses the Kamada-Kawai algorithm to search for the layout which least stretches or compresses the links between nodes. To reduce clutter only links between the first author and the others are shown, however neato treats *all* links as if they were the same length.

We construct a 942 × 942 symmetric integer matrix C. Where $C_{ij} = C_{ji}$ is the number of GP entries linking author i with author j and C_{ii} is the total number of GP

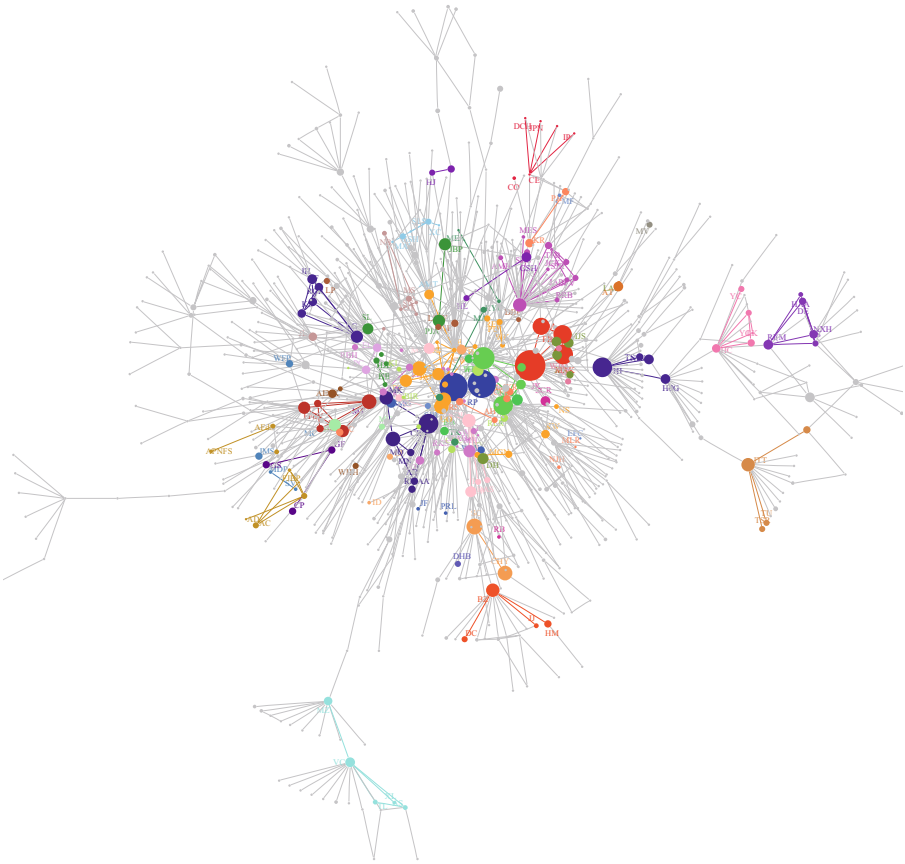

Fig. 1 Nine-hundred and forty-two people who have written one or more GP entries together. The area of circles is proportional to total number of entries in the GP bibliography. Lines indicate coauthorships. The largest components of the first 43 eigenvectors are coloured (using the same colours as in Table 1)

entries by author *i*. As expected, even in the largest connected component, most GP authors have not collaborated with most others, hence the elements of *C* are mostly zero, see Figs. 2 and 3.

Since *C* does not contain complex numbers and is symmetric, its eigenvalues and eigenvectors are also not complex. Positive eigenvalues correspond to clusters in the graph (non-zero blocks near the matrix diagonal) whilst negative eigenvalues correspond to partitioning the graph (off-diagonal blocks). The 942 eigenvalues of *C* are plotted in reverse order in Fig. 4. That is smallest eigenvalue on the right. 223 eigenvalues are within rounding error of zero. A further 264 are between zero and 1.0. That is half (487) of the eigenvalues are 1.0 or less. Figure 4 (solid line) shows the size of the eigen values grows rapidly. Indeed it (dashed line) suggests a power law relationship between size and rank.

Note the eigenvalues and eigenvectors form an orthonormal basis from which it is possible to exactly reconstruct the original matrix. The absolute magnitude of the eigenvalues gives the relative importance of each eigenvector. That is, it is possible to approximate the original matrix by ignoring eigenvectors whose eigenvalue is near zero. Another way of looking at this is to say that the large eigenvalue/eigenvectors pairs provide a lossy way to compress the matrix or capture its essence. As more eigenvectors are used the reconstruction becomes more accurate. Improvements can be either up or down. Positive eigenvector/eigenvalues increase matrix values whilst negative eigenvector/eigenvalues reduce them. We will come across reductions in Sect. 3.

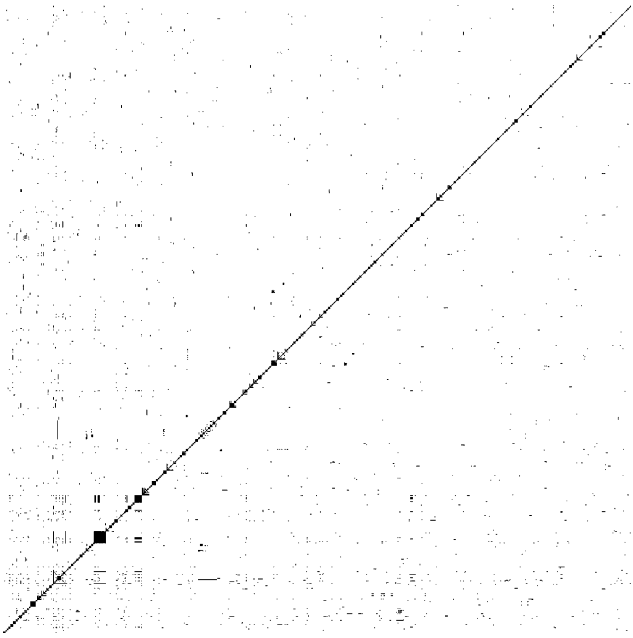

Fig. 2 Connectivity of giant component of the GP-bib connectivity matrix *C*. Squares on the diagonal are cliques formed by entries with a large number of coauthors, e.g. GECCO 2005 and GECCO 2003

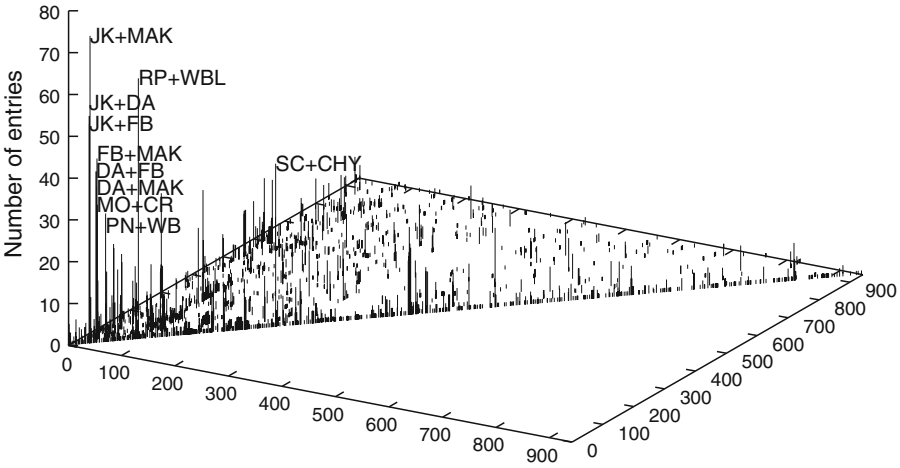

Fig. 3 Number of coauthored/coedited GP papers in the giant connected component of the GP-bib by coauthor pairings, cf. Fig. 2. Prolific pairings are annotated with the coauthors' initials. To avoid clutter, only collaborations are shown, i.e. the diagonal and the lower part of C are not plotted

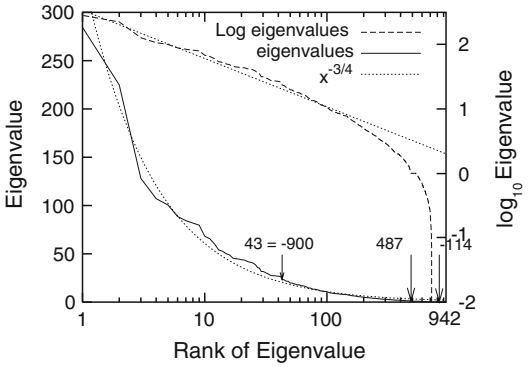

Fig. 4 Nine-hundred and forty-two Eigen values of the GP-bib connectivity matrix C

Figure 4 shows the size of eigenvalues falls rapidly. We imposed a cut off on the distribution and chose to study in detail the largest 43 eigenvalues and their eigenvectors. (i.e. leaving out the smaller 900 or so eigen components). Figure 5 shows the error in the matrix reconstructed from its eigen components falls exponentially (dotted line) as more components are used. With 43 eigenvectors 70% of the graph can be reconstructed and a 100% accurate reconstruction needs only 602 eigenvectors.

Remember the 942 eigenvectors form an orthonormal set. That is each is at right angles to the others by construction. Therefore, the elements of each eigenvector can be either positive or negative. For convenience they are normalised to be of unit length. We consider the largest (absolute) components of each of the 43 vectors so that together they account for 90% of the vectors' length. The number of components needed depends upon the vector and varies from 1 to 209. (From 2 to

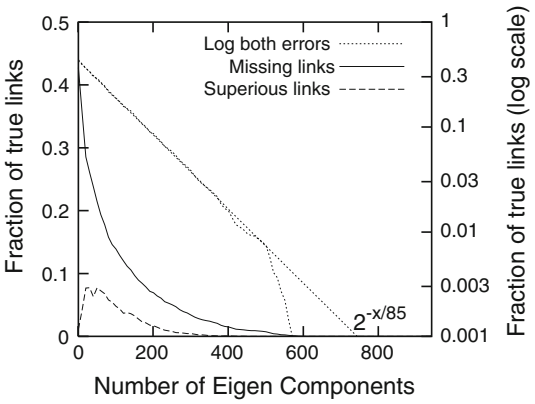

Fig. 5 Accuracy of reconstructed giant component of GP coauthor graph

60 in those corresponding to the 43 largest eigenvectors.) These are listed in Table 1.

Since Table 1 holds more than 90% of the data from the eigenvectors with the largest eigenvalues, we can say it captures the essential data from the connected component of the genetic programming bibliography. Of course this does not mean we could reconstruct the bibliography, or even the coauthor graph, from it. It does not contain fine enough detail for that. However, it does provide a summary of the coauthor relationships within genetic programming.

3 Analysis of eigenvectors

Table 1 contains eigenvectors sorted by order of their eigenvalue (first column). Each vector contains 942 elements, one per author. In Table 1 these are sorted by their absolute magnitude and only the name of the element and the values of the larger elements are given. Column 2 gives the number of elements displayed which, as described above, is the smallest number of elements needed to convey 90% of the eigenvector.

The first two eigenvectors are easy to interpret, cf. Fig. 6. All their large elements are positive and there are few of them. They summarise parts of the coauthor graph whose nodes (authors) do indeed collaborate and have many joint publications. However these provide only a summary of the interactions in both groups, as over time collaborative links might have been both strengthened and weakened. A single eigenvector cannot capture all of this information. Instead, as we shall see later, other eigenvectors strengthen and weaken links established earlier, so that the collaborations are represented with greater precision.

If we start at the top of Table 1 we can reconstruct our field in increasing detail. Each time we consider a new eigenvalue/eigenvector pair the resolution is improved.

The third eigenvector (Fig. 6) has four positive elements, corresponding to four collaborators but also a negative element (RP −0.17). The negative element does

Table 1 The largest 43 of the 942 eigenvalues of the largest connected component of the GP coauthorship graph. Eigenvalue, followed by the smallest number of components of the corresponding eigenvector which account for 90% of the vectors length, followed by the those components. Initials used to save space

284.55	4	John Koza 0.77, MAK 0.41, FB 0.33, DA 0.32
224.77	2	Riccardo Poli 0.73, WBL 0.66
128.08	5	Wolfgang Banzhaf 0.77, PN 0.54, FDF 0.18, RP −0.17, Markus Brameier 0.11
107.20	8	Conor Ryan −0.80, MO −0.46, MK −0.17, WBL −0.14, RP 0.13, RMAA −0.10, MN −0.10, AB −0.10
100.47	8	William B Langdon −0.71, RP 0.60, CR 0.17, NFM 0.14, WB 0.10, MO 0.10, JER 0.10, BB −0.09
88.71	2	Shu-Heng Chen −0.74, CHY −0.66
84.36	4	Hitoshi Iba −0.94, HdG −0.18, TS −0.16, NIN −0.16
81.83	5	Erik Goodman 0.55, JH 0.47, KS 0.42, RCR 0.39, ZF 0.36
79.71	5	Marco Tomassini 0.71, LV 0.48, FFdV 0.41, Philippe Collard 0.17, MC 0.15
68.11	12	David Andre 0.52, JK −0.52, FB 0.49, MAK 0.24, PN 0.22, WB −0.14, MJS −0.09, JPR −0.09, WJM −0.08, JY −0.08, HI 0.08, DH −0.06
65.51	25	Peter Nordin 0.68, WB −0.49, KW 0.19, DA −0.17, JK 0.16, FB −0.16, MGN 0.15, JAF −0.14, UO −0.11, JFM −0.10, TS −0.10, Markus Brameier −0.09, PD −0.09, FDF 0.07, MAK −0.06, EL 0.06, EC −0.06, NS 0.06, REK −0.06, WK −0.06, KD −0.05, HB −0.05, AMT −0.05, DD −0.05, LS −0.04
60.22	32	Una-May O'Reilly −0.48, JAF −0.38, JFM −0.38, TS −0.27, PN −0.20, WB 0.19, EC −0.18, HB −0.15, KD −0.14, DD −0.14, GK −0.13, RR −0.12, ACS −0.12, Mitchell A Potter −0.12, JW −0.12, RKS −0.12, Mark Harman −0.10, SWW −0.10, LD −0.10, KAD −0.10, Natasa Jonoska −0.10, FO −0.10, LS −0.10, TY −0.09, AMT −0.07, KW −0.06, RBH −0.06, MH −0.06, EB −0.06, WBL 0.05, MGN −0.05, DEG −0.05
54.07	10	Jason M Daida 0.84, JFV 0.24, TFB 0.24, SAS 0.19, SJR 0.18, RRB 0.16, JAP2 0.12, AMH 0.09, JCK 0.07, MES 0.07
52.70	6	Maarten Keijzer 0.72, VB 0.58, LS 0.20, CR −0.16, MC 0.12, JFM −0.11
50.02	18	Lee Spector 0.83, SL 0.21, MK −0.16, HJB 0.16, VB −0.15, HB 0.14, JY −0.12, MAK −0.12, WJM −0.12, DA 0.12, JAF −0.12, TS −0.11, JFM −0.11, PJA 0.10, UO 0.10, Jon Klein 0.09, Alan Robinson 0.07, JBP 0.05
48.69	10	Martin A Keane −0.48, JY −0.45, WJM −0.44, DA 0.38, JK 0.27, LS −0.22, GL −0.17, MJS −0.15, DH 0.11, OS −0.10
45.52	10	Jordan B Pollack 0.70, GSH 0.37, JFM −0.29, Pablo J Funes 0.24, PJA 0.21, TS 0.20, HL 0.19, HJ 0.17, JAF 0.14, TY −0.10
44.81	22	Terence Soule −0.53, JFM 0.45, JAF −0.42, JBP 0.31, TY 0.24, GSH 0.17, RBH −0.14, UO 0.11, Pablo J Funes 0.11, PJA 0.10, LS −0.09, KI −0.09, HJ 0.08, CDC 0.07, PT 0.06, TCF 0.06, HL 0.06, GK 0.05, EB 0.05, BIR −0.05, RLR 0.05, John Dickinson −0.04
44.52	4	Ivan T Tanev −0.90, TU −0.25, Katsunori Shimohara −0.23, TSR −0.16
43.80	4	Byoung-Tak Zhang −0.91, HM −0.29, DC −0.17, JJ −0.12
43.66	15	Thomas D Haynes −0.65, UO −0.44, JFM 0.32, SS −0.28, RLW −0.21, LS 0.16, DAS −0.16, FO −0.14, JAF 0.11, JBP 0.10, TS 0.08, MH −0.07, JR −0.06, MS 0.06, SL 0.06
42.62	20	Thomas D Haynes 0.55, UO −0.53, JFM 0.33, SS 0.26, RLW 0.19, FO −0.18, JAF 0.15, DAS 0.14, TS 0.12, TY −0.10, MS 0.10, JBP 0.09, LS 0.09, MH −0.09, EL 0.08, DEG −0.07, PC 0.07, RLR −0.06, EB 0.06, TCF 0.06
41.50	4	Daniel Howard 0.73, SCR 0.62, MO −0.17, RB 0.10
41.10	5	R I (Bob) McKay −0.68, NXH −0.45, DE −0.43, YS −0.23, HAA −0.22
40.29	15	Marc Schoenauer 0.63, EL 0.44, PC 0.42, JL 0.25, Frederic Raynal 0.16, JFM −0.15, UO 0.12, NB 0.11, OT 0.09, SG 0.09, MS 0.09, VB −0.07, TCF 0.07, TY −0.06, Habibou Maitournam 0.06
38.81	21	Tina Yu 0.79, CDC 0.31, TS 0.21, SMG −0.18, GK −0.15, EB −0.14, WPW 0.14, RLR 0.13, PRL 0.11, JF 0.10, JAF 0.10, UO −0.10, CHY −0.07, EC −0.07, RR −0.07, ACS −0.06, Mitchell A Potter −0.06, JW −0.06, MS 0.06, RKS −0.06, RBH 0.06
35.88	12	Steven M Gustafson −0.64, EB −0.44, GK −0.35, JFM 0.26, TY −0.24, NK −0.16, AE −0.13, WHH −0.11, CDC −0.11, UO 0.10, TCF 0.09, AMT 0.06
35.16	8	Hong-Qing Cao −0.59, LK −0.58, YC −0.34, Jingxian Yu −0.33, HdG −0.13, YC −0.10, NXH 0.10, RI(M 0.10
31.73	4	Giandomenico Spezzano −0.58, GF −0.56, CP −0.53, FFdV 0.18
31.64	13	Michael O'Neill 0.71, CR −0.41, AB 0.36, RMAA −0.18, MN −0.16, LI −0.14, TCF 0.13, DH 0.13, SCR 0.11, ID 0.10, JCW 0.09, GS −0.07, GF −0.07
30.93	20	Terence C Fogarty 0.73, JCW 0.54, MA 0.12, TK 0.11, MO −0.11, LB 0.10, SMG 0.10, EL −0.09, NFM −0.09, KYC 0.08, TS 0.08, PC −0.07, MEA 0.07, CR 0.07, JL −0.07, UO 0.06, AB −0.06, EB 0.05, JFM −0.05, AMT −0.05
29.97	11	Francisco Fernandez de Vega 0.55, Philippe Collard −0.46, MC −0.41, LV −0.31, JMS 0.30, GS 0.14, GF 0.13, CP 0.12, SV −0.11, MDP −0.10, WFP 0.09
29.27	13	Peter John Angeline −0.88, GSH 0.21, SL 0.19, DBF −0.13, LDW −0.11, HL 0.10, AMT 0.10, Pablo J Funes 0.10, JBP 0.09, JFM −0.09, GE −0.08, LP 0.07, TS −0.07
28.25	20	Andrew M Tyrrell −0.81, MAL −0.34, SLS −0.29, SMG −0.10, JAF 0.09, XC −0.08, GSH −0.08, PJA −0.08, RKS 0.08, Rodney Fry −0.07, RR 0.07, Mitchell A Potter 0.07, ACS 0.07, JW 0.07, Heng Liu −0.06, Mark Harman 0.06, LD 0.06, SWW 0.06, Natasa Jonoska 0.06, KAD 0.06
27.52	39	Justinian Rosca 0.63, SL 0.35, DHB 0.26, JFM 0.23, TS 0.18, MJS 0.16, LP 0.14, SMG 0.14, EC −0.14, KD −0.12, HB −0.12, UO 0.11, RKS −0.11, DD −0.10, MAK 0.10, RR −0.10, Mitchell A Potter −0.09, ACS −0.09, JW −0.09, FO 0.09, Mark Harman −0.08, AMT −0.08, FB −0.08, KAD −0.08, Natasa Jonoska −0.08, SWW −0.08, LD −0.08, LS −0.08, WJM −0.08, JY −0.07, TCF −0.06, RBH 0.06, JCW −0.06, EB 0.05, PT 0.05, TY −0.05, JAF −0.05, AE 0.04, CDC −0.04
27.36	10	Sean Luke −0.74, JR 0.43, LP −0.31, LS 0.18, DHB 0.18, PJA −0.17, MJS −0.12, HJB 0.09, HB 0.08, MAK −0.08
27.32	13	Matthew J Streeter 0.62, MAK 0.40, FB −0.29, WJM −0.29, JY −0.26, SL −0.26, JR −0.19, OS −0.14, NFM 0.13, LP −0.11, AT 0.09, LAB 0.09, DHB −0.08
27.12	60	Nicholas Freitag McPhee 0.43, JR −0.34, JFM 0.26, JER 0.21, TS 0.21, MJS −0.19, UO 0.17, PJA −0.16, EC −0.16, FO 0.15, DHB −0.14, KD −0.13, HB −0.13, MAK −0.13, AMT −0.13, SMG 0.13, RKS −0.12, RR −0.12, DD −0.11, ACS −0.11, Mitchell A Potter −0.11, JW −0.11, AHW 0.10, Mark Harman −0.10, SWW −0.09, Natasa Jonoska −0.09, KAD −0.09, LD −0.09, WJM 0.09, RP −0.09, KR −0.09, FB 0.08, NJH 0.08, JY 0.08, SL −0.08, TY −0.07, JAF −0.07, RBH 0.07, JCFP −0.06, CRS 0.06, PJF −0.06, EB 0.06, PT 0.06, MAL −0.06, MH 0.05, Ernesto Costa 0.05, CDC −0.05, AMMKR −0.05, SLS −0.05, GK −0.05, MO 0.04, GSH 0.04, LDW −0.04, OS 0.04, MLR 0.04, BB 0.04, LB −0.04, DBF −0.04, AE 0.04, MS −0.04
26.91	40	Nicholas Freitag McPhee 0.63, JR 0.33, JER 0.30, KR −0.20, JFM −0.16, AHW 0.15, PJF −0.14, DHB 0.14, RP −0.13, TS −0.13, SMG −0.12, NJH 0.12, JCW 0.10, UO −0.10, TCF 0.10, EC 0.10, JCFP −0.10, KD 0.09, CRS 0.09, FO −0.09, HB 0.09, RKS 0.08, AMMKR −0.08, DD 0.07, BB 0.07, RR 0.07, Mitchell A Potter 0.07, ACS 0.07, JW 0.07, SL 0.07, Mark Harman 0.06, SWW 0.06, LD 0.06, KAD 0.06, Natasa Jonoska 0.06, MLR 0.06, JAF 0.05, EFC 0.05, CMF −0.04, SJB 0.04
26.84	3	Astro Teller −0.87, MV −0.44, WJM −0.12
26.49	10	Katya Rodriguez-Vazquez 0.69, PJF 0.51, NFM 0.24, CMF 0.16, JPN 0.15, D Rees 0.15, IP 0.15, CE 0.15, DCH 0.15, CO 0.12
26.15	5	Victor Ciesielski 0.90, AS 0.23, MZ 0.19, TL 0.18, XL 0.12
23.51	14	Ernesto Costa −0.57, Penousal Machado −0.42, AC −0.42, FJBP −0.31, MS −0.17, AFdS −0.16, Jorge Tavares −0.16, JL 0.16, APNFS −0.15, EL 0.12, Sara Silva −0.09, PC 0.09, AD −0.08, VB 0.08

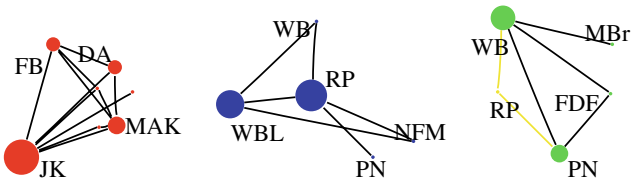

Fig. 6 Picture of first (left) second (center) and third (right) eigen components. To reduce clutter links with fewer than ten papers are not shown. Yellow indicates reduction in link strength

not mean that this author does not collaborate with the others. When we add the third eigenvector, we add four nodes (and up to 28 new links) but by the time of the third eigenvector there is overlap with the two larger eigen components. The negative value indicates the third eigenvector represents a tight group of four collaborators and that the collaborative links with one of the existing nodes in the reconstructed network must be weakened.

The fourth eigen component, is an example where most of the eigenvector's large components are negative. However there is one element (RP 0.13) with a large positive value, indicating that the overlap between this group of co-authors and the existing graph must be reduced for this element, rather than increased.

Most of the large elements of the fifth eigenvalue are positive whilst the largest is negative. Perhaps this can be thought of best, not as collaborators of the first author (WBL −0.71) but of the second (RP 0.60), where, again, the sign reversal indicates not that these two coauthors do not collaborate but that the links already established by the second eigenvector must be adjusted. Note one author (BB −0.09) has not collaborated with the second author but with the first and so his eigenvector element has the same sign as the first author and the opposite of the second.

The next four eigenvalues all have large elements of the same sign. Each represents an important collaboration in our field.

The tenth eigenvector (David Andre) is again of mixed sign and represents an important readjustment of the collaborative links. Most of the large elements represent adjustments to links between John Koza's coauthors and the addition of five of his coauthors who were not among the major component of the first eigenvector. Secondly, elements PN 0.22, WB −0.14 and HI −.08 adjust links between coauthors in the third and seventh eigenvectors.

The eleventh eigenvector (Peter Nordin) can be thought of as similar to the fifth. It, too, contains a mixture of positive (PN) and negative (WB) values. The positive elements creating links between PN and important collaborators, whilst the negative ones mostly create new links for WB. While others appear to be making modest adjustments to existing members of the graph. Note some links are established by collaborations on coediting conference proceedings.

The twelfth eigenvector is the first one to contain many large elements (32). Many of these (which are all of the same sign) represent collaborative links established by Una-May O'Reilly through coediting Advances in Genetic

Programming, EuroGP, GECCO and Genetic Programming Theory and Practice. Others represent collaborations on papers.

The next eigenvector has all positive large elements and introduces Jason Daida and his collaborators. The fourteenth summarises the collaborations of Maarten Keijzer (note the data are prior to the publication of GECCO 2006, which he edited). However, two opposite sign elements (CR −0.16 and JFM −0.11) again indicate adjustment to existing links. The fifteenth eigenvector also contains many large elements (18). These are mostly the collaborators of Lee Spector. Again elements with the opposite sign suggest adjustment to the existing network (particularly for John Koza's coauthors and those introduced by the previous eigenvector). While one element (JBP 0.05) represents the addition of a coauthor of a coeditor (PJA 0.10) of Lee.

Ten of the remaining 28 large eigen components have relatively few large elements and these are mostly of the same sign. They can be thought of as mostly describing an important group of collaborating authors. Other eigenvectors tend to have more large components with more evenly matched number of positive and negative elements. While they introduce new authors they also play an important role in adjusting strengths of existing links.

The two eigenvectors for Tom Hayes have similar eigenvalues (≈ 43) and are relatively diffuse (15 and 20 major elements) but 12 authors play major roles in both. Again positive and negative values appear, indicating these two eigen components play an important role in recording the asymmetry in the graph. Similarly the pair of eigenvectors for Nick McPhee (60 and 40 components) both have eigenvalues of about 27 and have 35 coauthors in common. The mix of signs indicates the formation of clusters with strong links within a cluster but no or weak links between strong centres.

4 Conclusions

Graphical tools, like neato, provide a valuable way to represent the coauthor links within the genetic programming bibliography. However, with large networks like this, one is rapidly swamped by details and it is difficult to annotate them meaningfully. Eigen analysis provides a ready way to extract quantitative signals from the data in a significant order.

As one studies Table 1 from the top, one quickly moves from anticipated relationships to surprises. These where discussed in Sect. 3.

Acknowledgements We are grateful for free use of neato (http://www.graphviz.org) and Octave (http://www.octave.org).

Appendix

Table 2 Initials used in Table 1 and Figs. 1, 3, 6

AB	Anthony Brabazon	AC	Amilcar Cardoso	ACS	Alan C Schultz	AD	Andre Dias	AE	Aniko Ekart
AFdS	Arlindo Ferreira da Silva	AHW	Alden H Wright	AMH	Adam M Hilss	AMM-KR	Amr Mohamed Mahmoud Khairat Radi	AMT	Andrew M Tyrrell
APN-FS	Ana Paula Neves F Silva	–	Alan Robinson	AS	Andy Song	AT	Astro Teller	BB	Bernard Buxton
BIR	Bart I Rylander	BZ	Byoung-Tak Zhang	CDC	Chris D Clack	CE	C Evans	CHY	Chia Hsuan Yeh
CMF	Carlos M Fonseca	CO	Carlos Oliver-Morales	CP	Clara Pizzuti	CR	Conor Ryan	CRS	Christopher R Stephens
DA	David Andre	DAS	Dale A Schoenefeld	DBF	David B Fogel	DC	Dong-Yeon Cho	DCH	D C Hill
DD	Dipankar Dasgupta	DE	Daryl Essam	DEG	David E Goldberg	DH	Daniel Howard	DHB	Dana H Ballard
–	D Rees	EB	Edmund Burke	EC	Erick Cantu-Paz	EFC	Ellery Fussell Crane	EG	Erik Goodman
EL	Eduard Lukschandl	FB	Forrest Bennett	FDF	Frank D Francone	FFdV	Francisco Fernandez de Vega	FJBP	Francisco Jose Baptista Pereira
FO	Franz Oppacher	–	Frederic Raynal	GE	Gusz Eiben	GF	Gianluigi Folino	GK	Graham Kendall
GL	Guido Lanza	GS	Giandomenico Spezzano	GSH	Gordon S Hollingworth	HAA	Hussein A Abbass	HB	Howard Barnum
HC	Hong-Qing Cao	HI	Hitoshi Iba	HJ	Hugues Juille	HJB	Herbert J Bernstein	HL	Hod Lipson
–	Heng Liu	HM	Heinz Muhlenbein	–	Habibou Maitournam	HdG	Hugo de Garis	ID	Ian Dempsey
IP	I Pratt	ITT	Ivan T Tanev	JAF	James A Foster	JAP2	John A Polito 2	JBP	Jordan B Pollack
JCFP	Joao Carlos Figueira Pujol	JCK	Jonathan C Khoo	JCW	James Cunha Werner	–	John Dickinson	JER	Jonathan E Rowe
JF	Jonny Farringdon	JFM	Julian F Miller	JFV	John F Vesecky	JH	Jianjun Hu	JJ	Je-Gun Joung
JK	John Koza	–	Jon Klein	JL	Jean Louchet	JMD	Jason M Daida	JMS	J M Sanchez
JPN	J P Norton	JPR	James P Rice	JR	Justinian Rosca	–	Jorge Tavares	JW	Joachim Wegener
JY	Jessen Yu	–	Jingxian Yu	KAD	Kathryn A Dowsland	KD	Kalyanmoy Deb	KI	Kosuke Imamura

Table 2 continued

KR	Katya Rodriguez-Vazquez	KS	Kisung Seo	–	Katsunori Shimohara	KW	Krister Wolff	KYC	Kit Yan Chan
LAB	Lee A Becker	LB	Larry Bull	LD	Lawrence Davis	LDW	L Darrell Whitley	LI	Laur Ivan
LK	Li-Shan Kang	LP	Liviu Panait	LS	Lee Spector	LV	Leonardo Vanneschi	MA	Manu Ahluwalia
MAK	Martin A Keane	MAL	Michael A Lones	–	Mitchell A Potter	–	Markus Brameier	MC	Mike Cattolico
MDP	Michael Defoin Platel	MEA	Mehmet Emin Aydin	MES	Michael E Samples	MGN	Mats G Nordahl	MH	Martin Hemberg
–	Mark Harman	MJS	Matthew J Streeter	MK	Maarten Keijzer	MLR	Mitchell L Reierson	MN	Miguel Nicolau
MO	Michael O'Neill	MS	Michele Sebag	MT	Marco Tomassini	MV	Manuela Veloso	MZ	Mengjie Zhang
NB	Nicolas Bredeche	NFM	Nicholas Freitag McPhee	NIN	Nikolai I Nikolaev	NJH	Nicholas J Hopper	–	Natasa Jonoska
NK	Natalio Krasnogor	NS	Nils Svangard	NXH	Nguyen Xuan Hoai	OS	Oscar Stiffelman	OT	Olivier Teytaud
PC	Pierre Collet	–	Philippe Collard	PD	Peter Dittrich	PJA	Peter John Angeline	PJF	Peter J Fleming
–	Pablo J Funes	–	Penousal Machado	PN	Peter Nordin	PRL	Peter R Lidwell	PT	Peter Thomson
RB	Richard Brankin	RBH	Robert B Heckendorn	RCR	Ronald C Rosenberg	REK	Robert E Keller	–	Rodney Fry
RI(M)	R I (Bob) McKay	RKS	Russell K Standish	RLR	Rick L Riolo	RLW	Roger L Wainwright	RMAA	R Muhammad Atif Azad
RP	Riccardo Poli	RR	Rajkumar Roy	RRB	Robert R Bertram	SAS	Stephen A Stanhope	SC	Shu-Heng Chen
SCR	Simon C Roberts	SG	Sylvain Gelly	SJB	S J Barrett	SJR	Steven J Ross	SL	Sean Luke
SLS	Stephen L Smith	SMG	Steven M Gustafson	SS	Sandip Sen	–	Sara Silva	SV	Sebastien Verel
SWW	Stewart W Wilson	TCF	Terence C Fogarty	TDH	Thomas D Haynes	TFB	Tommaso F Bersano-Begey	TK	Tatiana Kalganova
TL	Thomas Loveard	TS	Taisuke Sato	TSR	Thomas S Ray	TU	Takashi Uozumi	TY	Tina Yu
UO	Una-May O'Reilly	VB	Vladan Babovic	VC	Victor Ciesielski	WB	Wolfgang Banzhaf	WBL	William B Langdon
WFP	William F Punch	WHH	William H Hsu	WJM	William J Mydlowec	WK	Wolfgang Kantschik	WPW	William P Worzel
XC	Xinye Cai	XL	Xiang Li	YC	Yongyan Chen	YS	Yin Shan	ZF	Zhun Fan

References

1. A.J. Lotka, The frequency distribution of scientific productivity. J. Wash. Acad. Sci. **16**(12), 317–323 (1926)
2. W. Shockley, On the statistics of individual variations of productivity in research laboratories. Proc. Inst. Radio Eng. **45**, 279–290 (1957)
3. S. Wassermann, K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994)
4. M.E.J. Newman, The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. U.S.A. **98**(2), 404–409 (2001)
5. M.E.J. Newman, Coauthorship networks and patterns of scientific collaboration. Proc. Natl. Acad. Sci. U.S.A. **101**(Suppl 1), 5200–5205 (2004)
6. M. Turk, A. Pentland, Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
7. B.F. Buxton, M.B. Dias, Implicit, view invariant, linear flexible shape modelling. Pattern Recogn. Lett. **26**(4), 433–447 (2005)
8. W.W. Sherrill, MD/MBA students: an analysis of medical student career choice. Med. Educ. Online **9** (2004)
9. J.E. Andrews, An author co-citation analysis of medical informatics. J. Med. Libr. Assoc. **91**(1), 47–56 (2003)
10. M. Tomassini, L. Luthi, M. Giacobini, W.B. Langdon, The structure of the genetic programming collaboration network. Gen. Prog. Evol. Mach. **8**(1), 97–103 (2007)
11. W.B. Langdon, S. Gustafson, Genetic programming and evolvable machines: Five years of reviews. Gen. Prog. Evol. Mach. **6**(2), 221–228 (2005)
12. S. Milgram, The small-world problem. Psychol Today **1**(1), 60–67 (1967)